

Internet et les moteurs de recherche

Qu'est ce qu'un site internet ? C'est une interface graphique permettant de lire en clair des informations écrites avec un code de programmation spécifique (appuyez sur F12 ou clic droit > inspecter pour un aperçu).



Pour exister le site web doit respecter ces 2 conditions :

1. Il a besoin d'un **serveur** contenant les **fichiers** et la **base de données**, dans le cas d'un site internet comme celui que vous êtes en train de lire la langue principale est le PHP et MySQL. C'est le serveur qui traduit et exécute les ordres d'envois en fonction de la demande de l'internaute.

2. Et d'un **Un navigateur web** compatible pour traduire le code et afficher le site. Navigateurs web connus : Google Chrome, Edge, Opera , Firefox

Pour passer du serveur distant à votre navigateur, on a besoin d'un **Nom de Domaine (ndd)** : c'est à dire une adresse physique compréhensible par l'humain. L'organisme de gestion des noms de domaine s'appelle l'ICANN. Un nom de domaine est la composition du nom + l'extension niicolab.com et elle est unique, mais il peut exister des niicolab.org ou niicolab.co.uk. Tout le monde peut acheter et déposer un nom de domaine.

L'adresse internet se présente sous la forme 172.217.4.206 on appelle ça une adresse IP. Mais comme ça serait trop complexe pour le commun du mortel d'apprendre toutes ces suites nombres, nous utilisons un format plus compréhensible L'URL. Et elle se compose de :

le nom du domaine (education.fr, niicolab.com, openoffice.org, .gouv.fr etc.) : il permet souvent de repérer la **nature** (universitaire, institutionnel, commercial, associatif, etc.) et le **statut** du site (officiel, privé, personnel, etc.) ;

l'extension du domaine caractérise l'orientation générale du site (langue .fr .it .de, commercial .com .net , organisation .org .gouv etc.) ;

le sous-domaine éventuel (indiqué en préfixe), ici: formation.niicolab.com ou formation est le sous domaine du site niicolab.com

les sous dossiers d'organisation du site <https://formation.niicolab.com/lecon/le-reseau> les "/" marque l'arborescence du site

Moteur de Recherche

Et pour s'y retrouver on utilise un **moteur de recherche** qui est un outil permettant de retrouver des pages web à partir d'une requête.

Les moteurs de recherches les plus courant sont Google, Bing, Yahoo, et d'autres moins connus Qwant, Duck Duck GO, Ecosia

il réalise pour vous les tâches suivantes :

Moissonnage (ou Crawling) : les pages du web sont parcourues automatiquement par un robot, et tout le contenu s'en retrouve indexé. C'est pour cette raison que quand vous faites une recherche vous pouvez tomber sur des documents, des photos ou des vidéos par exemple (en cliquant sur l'onglet dédié)

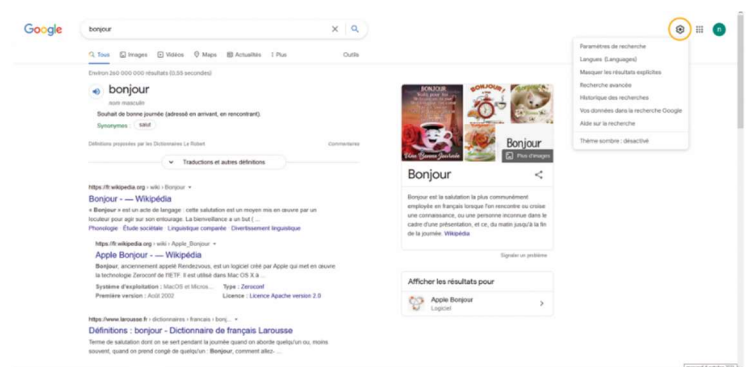
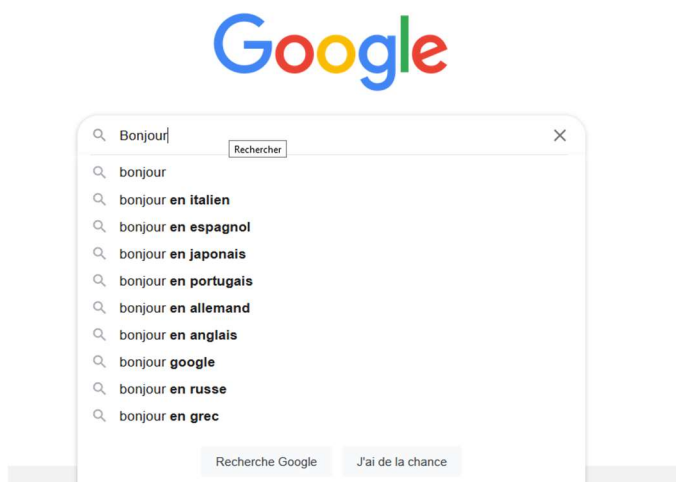
indexation automatique : les pages sont analysées pour en extraire les mots clés, à des fins de classification. Pour indexer les images, les moteurs utilisent les métadonnées associées, le texte entourant l'image ou les tags laissés par les internautes. Le créateur du site à un très gros travail à faire pour que l'indexation soit propre, cela s'appelle le Référencement.

Interroge sa base de donnée selon votre requête : les mots clés de la requête sont comparés aux mots extraits par l'indexation et une liste de pages web sélectionnées est affichée par ordre de pertinence.

Adapte les résultats à votre **localisation** Selon le lieux ou vous vous trouvez, les réponses vont être géographiques et ne donneront pas forcément les mêmes résultats selon la region ou la langue. Une recherche en français au Québec va d'abord chercher l'information ...au Québec puis va élargir au web francophone. Pratique si on recherche un restaurant par exemple.

Il existe des **moteurs de recherche spécialisés** dans certains types de ressources (images, vidéos, publications académiques, blogs, flux RSS, personnes, etc.) ou dans certains domaines (technologie, musique, mathématiques, santé, chimie, etc.). *Sur certains sites web, on trouve un moteur dédié aux pages du site lui-même.*

Recherche Bonjour



Exemple de résultat de la recherche (le 6/10/2021)

La **recherche d'information** nécessite la mise en place d'une stratégie en 4 étapes qui consiste à :

- Cerner et formuler le besoin d'information ;
- Lire les résultats et repérer les sources pertinentes ;
- identifier les outils à exploiter en fonction de ces sources ;
- Quelle utilisation je peux en faire

Cerner et formuler la recherche pour un résultat pertinent en utilisant un moteur de recherche

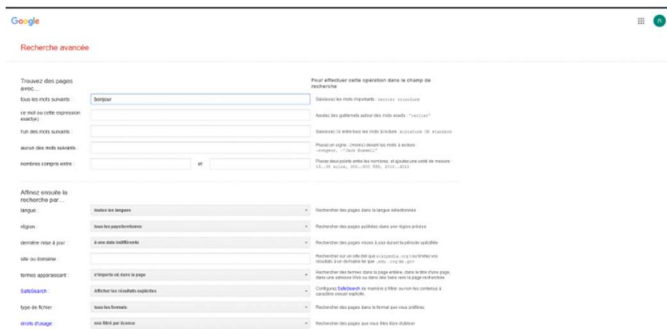
Les moteurs de recherche tendent à limiter l'effort de formulation de l'utilisateur en décidant automatiquement de la façon de combiner et ajuster les mots-clés saisis. Cependant, il est possible de réaliser une recherche plus fine en utilisant les opérateurs booléens pour combiner les critères.

Les moteurs proposent généralement des notations plus légères :

- le signe + devant un mot-clé signifie qu'il est obligatoire (proche du ET) ;
- le signe - signifie qu'il ne doit pas apparaître (proche du SAUF) ;

- les **guillemets** pour rechercher une expression exacte (l'expression entière doit être présente, et sous cette forme exactement).
- Des codes comme Filetype: ou Site: resserrent encore plus la recherche.

La page de **recherche avancée** de google (entouré en Orange sur la photo exemple de résultats à gauche)



Les moteurs offrent souvent un mode de "recherche avancée" où on peut préciser par exemple la langue des réponses, les sites auxquels doivent se limiter les réponses, etc.

Vous recentrez votre recherche.

L'**ordre des réponses** d'un moteur à une requête tient compte de la **pertinence thématique**, c'est-à-dire de l'importance des mots-clés de la requête dans la page web. Il tient aussi compte de la **popularité** de la page web, que les moteurs déduisent :

de l'abondance et de la qualité des liens qui mènent à la page ;

de la tendance des internautes à cliquer sur la page en réponse à la même requête

Lire les résultats des recherches, évaluer les sources

Où est ce que je trouve l'information ? Et qui l'a écrite

Parmi les sources générales d'informations, qu'on retrouve le plus souvent sur le web :

- **les sources d'informations traditionnelles** (agences de presse, grandes organisations d'intérêt privé ou public, associations d'experts, publications scientifiques, etc.) qui diffusent des informations validées selon un processus admis par eux-mêmes et leurs pairs; *l'organisation internationale OCDE, la Bibliothèque Nationale de France, les éditeurs de publications scientifiques).*
- **les sources de structures collectives plus ou moins formelles**, dont les règles de publication et de validation de l'information sont plus ou moins strictes. *L'encyclopédie en ligne Wikipédia est une source qui émerge d'une communauté d'auteurs ; la qualité de l'information qu'elle diffuse est liée aux règles et processus d'édition qui lui sont propres, à lire avec précaution selon le sujet de la recherche.*
- **les sources informelles** publiées sans aucun contrôle. les pages blogs de site de journaux, les site personnels.
- Le **journalisme** est un cas à part, car il subsiste 2 formes principales, il faut croiser les données
 - Les grands journaux nationaux sont des sources d'information fiables généralement admises par tout le monde. Mais, attention, les informations publiées sont très dépendantes de l'orientation politiques et des financeurs externes. (BFTMtv, LCI, lemonde, l'AFP, l'obs, le figaro ou Marianne par exemple)
 - Les journaux indépendant sont des sources d'analyses d'informations fiables, mais n'ont généralement pas les moyens financiers pour assurer une mise à jour quotidienne, et ont un pouvoir de diffusion moindre (uniquement web par exemple)
- **Les réseaux sociaux** sont une jungle où toutes les sources d'informations se chevauchent, se valident et se contredisent à la fois.

Et donc il faut bien faire attention à l'auteur (ou celui qui publie l'info)

- Son **activité** professionnelle ou associative (métier, expérience, réalisations, statut, niveau d'études)
- Son **affiliation et déclaration d'intérêt** (entreprise, organisation, sponsor, partenaires, gouvernement, université, etc.) ;
- Son **bord** politique (le trotskisme n'a pas la même "définition" à gauche et à droite)
- Sa **reconnaissance** par ses pairs.
- Son "historique" de **fiabilité** : c'est-à-dire la confiance qu'on peut accorder à la **véridicité** ses écrits (a t'il déjà trompé volontairement ou non)

Identifier le public visé et les types de publications

Le public visé (grand public, spécialistes) contribue à éclairer le but dans lequel elle à été produite.

La fiabilité d'une information est liée au contexte dans lequel elle a été communiquée (processus de validation) et à l'intention sous-jacente (commerciale, académique, officielle, etc.).

La fraîcheur de l'information, il faut repérer la date de création de la ressource et éventuellement la date de sa dernière mise à jour.

Les types de publications peuvent être généralistes ou spécialisées. On peut les organiser ainsi :

- Manuels fondamentaux (écrits par les auteurs fondateurs d'un domaine) et ressources pédagogiques, sites d'universités
- Encyclopédies, dictionnaires et lexiques (édités par les éditeurs importants d'un domaine)
- Revues scientifiques (souvent éditées par des sociétés savantes, et dont les articles sont validés par des experts) et littérature grise (rapports d'étude ou de recherche, actes de congrès, thèses, brevets, etc.) ;
- Vulgarisation scientifique (souvent sous la forme de revues ou d'ouvrages destinés au grand public) ;
- Actualité (presse quotidienne tournée vers le grand public ou un lectorat plus spécialisé selon le cas) ;
- Rapports et bilans (produits par des organisations et institutions reconnues au niveau national ou international)
- sources officielles (produites par les gouvernements ou organismes gouvernementaux) ;
- Banque de données et de statistiques (produites par des groupements de chercheurs ou des organismes officiels) ;
- Blogs et podcasts (édités par des internautes relevant ou non d'institution bien identifiée).

Lorsqu'on se documente sur un domaine que l'on connaît mal, il faut commencer par repérer les sources pertinentes qui permettront de cerner les concepts et la terminologie associés : dictionnaires, lexiques, encyclopédies, manuels, etc.

Attention à l'angle de lecture d'une information sur internet « l'information connaissance » VS « l'information d'opinion » (information journalistique, publicitaire, issue des pseudo-sciences, qui est le produit du sens commun, de l'évidence, des préjugés, des stéréotypes). L'OMC ne dira pas les même choses sur le Capitalisme que MSF mais les résultats des études seront sensiblement les mêmes.

La qualité des sources vs les infox (fausses informations), et théorie du complot.

Pour évaluer la qualité d'une source, on pourra s'intéresser à sa fiabilité (propension à diffuser des informations exactes, liée au processus de validation de l'information) et à son autorité (réputation dont elle jouit dans son domaine auprès des experts).



La définition Wikipédia de fake news, en français infox, informations fallacieuses, fausses informations ou fausses nouvelles, constituent une information mensongère délivrée dans le but de manipuler ou tromper un auditoire. Prenant une importance singulière à l'ère d'internet, elles peuvent émaner d'un ou plusieurs individus, d'un ou de plusieurs médias, d'un homme d'État ou d'un gouvernement.

- La terre est plate n'est n'est pas du complotisme, c'est une fausse informations.
- A ne pas confondre fakenews et complotisme.

la théorie du complot, également désignée par les néologismes conspirationnisme ou complotisme propose de donner une vision de l'histoire perçue comme le produit de l'action d'un groupe occulte agissant dans l'ombre.

C'est à vous

Qu'elle est la dernière recherche que vous ayez faites ?

Sur quel moteur de recherche ?

Sur quels sites allez-vous habituellement ?

Ajoutez vos notes
